A Correlation Analysis of English Particle Placement of Three East Asian EFL Learners' Writings

Ha-Eung Kim
Hannam University
70 Hannamro, Daedeok-gu
Daejeon 306-791, Korea
tankkh@hanmail.net

Gyu-Hyeong Lee
Hannam University
70 Hannamro, Daedeok-gu
Daejeon 306-791, Korea
gyuhyung73@naver.com

Yong-hun Lee
Chungnam Nat'l University
99 Daehak-ro, Yuseong-gu
Daejeon 305-764, Korea
yleeuiuc@hanmail.net

Abstract

This paper examines the English particle placements of EFL learners' writings in three East Asian countries (Chinese, Japan, and Korea). Three parts of the TOEFL11 corpus were chosen, and all the sentences with particles were extracted. The ICE-GB was chosen as a native speakers' English. Then, eleven linguistic factors were manually encoded. The collected data were analyzed with R. Correlation tests and a hierarchical clustering analysis was adopted. Through the analysis, the following two facts were observed: (i) each linguistic factor affected differently in four varieties of English and (ii) Japanese English was similar to native speakers' counterparts whereas Korean and Chinese formed another group.

1 Introduction

Linguistic alternation has been one of the interesting research areas in linguistics. Particle placement is one of such syntactic alternations. It refers to the linguistic phenomenon where a particle is located before or after the direct object (DO) in the phrasal verb constructions.[1] For example, let's see the following sentence (Gries, 1999:1).

(1)   a. John *picked up* the book.
       b. John *picked* the book *up*.

As you can see, the word order in (1a) is 'verb + particle + DO', whereas the order of (1b) is 'verb + DO + particle'.

There have been a lot of studies on this topic in traditional grammar and Chomskyan syntax. They have primarily focused on what linguistic factors determine the choice of alternations. Nowadays, as computer technology and statistics develop, there have been a few corpus-based studies to explain these syntactic phenomena with authentic corpus data and statistical analysis. Gries (1999, 2001, 2003) were such trials, and these studies adopted a multifactorial analysis to investigate the particle placement in the native speakers' writings. These studies also proposed several linguistics factors and the factors were encoded in the corpus data. These studies demonstrated that various linguistic factors and their interactions with the main factors significantly influenced the choice of alternations

This paper, however, adopted a monofactorial analysis to examine the particle placement in three Eat Asian EFL learners' writings (Korean, Chinese, and Japanese). The TOEFL11 corpus was used for the EFL learners' writings, and the ICE-GB corpus (the British component of the International Corpus of English; Nelson et al., 2002) was chosen for the native speakers' counterparts.

[1] Gries (1999) used the term *particle movement* while Gries (2001) used the term *particle placement*. The former adopted Chomsky's transformational-generative grammar approach (Chomsky, 1957, 1965) and thought that particle moved from one position to another. The latter did not presuppose such movement analysis. This paper adopted Gries' second approach and called the phenomena in (1) *particle placement*. That is, this paper did not presuppose the movement of particles. Instead, how various linguistic factors influenced the placement of particles was investigated with statistical tools.

30th Pacific Asia Conference on Language, Information and Computation (PACLIC 30)
Seoul, Republic of Korea, October 28-30, 2016

347

From these four corpora (Chinese, Japanese, Korean, and ICE-GB), all the relevant sentences were extracted using the tag information. Then, eleven linguistic factors were manually encoded to these sentences. After the process, all the linguistic factors were statistically analyzed with R. Two different types of statistical analyses were adopted in the paper: correlation analysis and a hierarchical clustering. These statistical analyses demonstrated how each linguistic factor played a role in the choice of particle placement, in the four varieties of English.

This paper is organized as follows. In Section 2, three groups of previous studies are reviewed with a focus on corpus-based approaches. Section 3 is on the corpus data and research methods. Section 4 contains the analysis results of correlation analyses, and Section 5 the analyses results of a hierarchical clustering. Section 6 is for discussions, and Section 7 summarizes this paper.

2 Previous Studies

2.1 Studies in Traditional Grammar

There have been several studies on English particle placement in various linguistic fields: traditional grammar (Sweet, 1892; Jespersen, 1928; Kruisinga and Erades, 1953), Chomskyan transformational-generative grammar (Fraser, 1974, 1976; Den Dikken, 1992, 1995; Rohrbacher, 1994), cognitive grammar (Yeagle, 1983), discourse-functional approaches (Chen, 1986), psycholinguistically-oriented approaches (Hawkins, 1994), and so on.

In the traditional grammar, there have been lots of studies on English particle placement (Sweet, 1892; Jespersen, 1928; Kruisinga and Erades, 1953). Gries (1999:33) closely investigated the claims in previous studies and summarized them as in Table 1.

Here, $construction_0$ refers to the sentences with the order of 'verb + particle + DO' as in (1a), while $construction_1$ refers to the sentences with the order of 'verb + DO + particle' as in (1b). Table 1 enumerated 18 different linguistic factors and this table demonstrated that several different types of factors, not a single factor, actually influenced the choice of alternations.

Value for $construction_0$	Variable	Value for $construction_1$
Long DO	Length of the DO in words (Length W)	
Long DO	Length of the DO in syllables (LengthS)	
Complex	Complexity of the DO (Complex)	
	NP-Type of the DO: semi-pronominal (Type)	pronominal
Indefinite	Determiner of the DO (Det)	definite
No	Previous mention of the DO (Lm)	yes
Low	◄──Times of preceding mention of the DO (Topm)──►	high
High	◄──Distance to last mention of the DO (Dtlm/ActPC)──►	low
High	◄──News Value of the DO──►	low
Yes	(Contrastive) Stress of the DO	
Yes	Subsequent mention of the DO (Nm)	no
High	◄──Times of subsequent mention of the DO (Tosm)──►	low
How	◄──Distance to next mention of the DO (Dtnm/ClusSC)──►	high
	Overall frequency of the DO (Om)	
	following directional adverbial (PP)	yes
Yes	Prep of the following PP is identical to the particle (Part = Prep)	
	Register	
Idiomatic	◄──Meaning of the VP (Idiomaticity)──►	literal
Low	◄──Cognitive Entrenchment of the DO──►	high
Inanimate	Animacy of the DO (Animacy)	animate
Abstract	Concreteness of the DO (Concreteness)	concrete

Table 1: Variables That Govern the Alternation

Let's examine how these factors can be related with the choice of particle placement. For example, LENGTHW (the first factor in Table 1) refers to the length of DO in words. If the DO is long, native speakers tend to choose $construction_0$ rather than $construction_1$. If the DO is short, however, the native speakers tend to use $construction_1$ rather tahn $construction_0$. The factor DET, the fifth factor, refers to the determiner of the DO. If the determiner of DO is indefinite (such as *a* or *an*), native speakers tend to choose $construction_0$ rather than $construction_1$. If the determiner of DO is definite (such as *the*), native speakers prefer to use $construction_1$ rather than $construction_0$. Table 1 contains all the related factors which cover most of linguistic fields: phonology, syntax, semantics, pragmatics, and discourse analysis.

2.2 Gries' Corpus-based Approaches

Though it is fact that previous studies in traditional grammar surely contributed to the study of particle placement, their data exclusively relied on native speakers' intuition. Gries (2001, 2003) pointed out this problem and performed an analysis based on the corpus data.

Gries (2001:36-37) pointed out the problems of these previous approaches, and Gries (2001, 2003) employed a corpus-based analysis. They adopted both monofactorial analyses and multifactorial analyses.

In the monofactorial analyses, each linguistic factor was statistically analyzed. In these studies, the British National Corpus (BNC; Aston and Burnard, 1998) was taken, and all the sentences

with phrasal verbs were extracted. Then, several linguistic factors were manually annotated, and the data were statistically analyzed. Two types of statistical analyses were taken. The first one is correlation analysis, and each factor was analyzed as follows (Gries, 2001:42).

Variable/Variable: *Value*	Correlation Coefficient
Complexity of the DO	γ =-0.85***
Idiomaticity of the VP	γ =-0.6***
Complex: *simple* NP	ϕ=0.522*** (λ=0.49)
NP Type of the DO	ϕ=0.492*** (λ=0.366)
Length of the direct object in syllables	r_{pbis}=-0.481 ***
Type: *lexical* NP	ϕ=0.47*** (λ=0.366)
Type: *pronominal* NP	ϕ=0.468*** (λ=0.32)
Complex: *intermediate* NP	ϕ=0.455*** (λ=0.412)
Distance to last mention of the DO	r_{pbis}=0.452 ***
Cohesiveness of the DO to the preceding discourse	r_{pbis}=0.429 ***
Length of the DO in words	r_{pbis}=0.423 ***
Times of preceding mention of the DO	r_{pbis}=0.414 ***
Last mention of the DO	ϕ=0.411*** (λ=0.387)
Overall mention of the DO	r_{pbis}=0.357 ***
Concreteness of the DO	ϕ=0.339*** (λ=0.314)
Idiomaticity: *idiomatic* VP	ϕ=-0.328*** (λ=0.253)
Determiner of the DO	ϕ=0.319*** (λ=0.206)
Idiomaticity: *literal* VP	ϕ=0.314*** (λ=0.268)
Register	ϕ=0.291*** (λ=0.263)
DET. *indefinite determiner*	ϕ=-0.288*** (λ=0.206)
Directional adverbial	following the DO
DET. *no determiner*	ϕ=0.232*** (λ=0.191)
Complex: *complex* NP	ϕ=-0.193*** (λ=0.077)
Times of subsequent mention of the DO	r_{pbis}=0.191 ***
Animacy of the DO	ϕ=0.166*** (λ=0.057)
Cohesiveness of the DO to the subsequent discourse	r_{pbis}=0.142 ***
Next mention of the DO	ϕ=0.104* (λ=0.072)
Distance to next mention of the DO	r_{pbis}=0.1*
Type: *semi-pronominal* NP	ϕ=0.092*** (λ=0)
Idiomaticity: *metaphorical* NP	ϕ==0.047 ns (λ=0.016)
Type: *proper name*	ϕ=0.023 ns (λ=0)
DET: *definite determiner*	ϕ=-0. 018 ns (λ=0)
Particle equals the preposition of the following PP	ϕ=0.003 ns (λ=0)

Table 2: Correlation Analysis in Gries (2001)

As you can observe in this table, Gries (2001) calculated the correlation coefficients for both each linguistic factor and each value for the factor. In addition to the correlation analysis, Gries (2001) also took a linear discriminant analysis (LDA), where factor loading of each factor was calculated.

Gries (2001, 2003) also employed multifactorial analyses, where all the linguistic factors in Table 1 were taken into consideration simultaneously. The studies used a Generalized Linear Model (GLM) to statistically analyze how each linguistic factor played a role in the choice of alternation. These studies also took a classification and regression tree (CART) and calculated the importance of each factor.

Gries (2001) and Gries (2003) were essentially different from the previous approaches, since (i) these studies were based on corpus data (naturally occurring data) and (ii) they statistically analyzed the collected data.

2.3 Lee et al. (2015) and Lee et al. (2016)

Following the analyses in Gries (2001, 2003), Lee et al. (2015) analyzed the particle placement in the Korean EFL learners' writings. The studies used the Korean component of the TOEFL11 corpus (which was the same corpus that this paper used) to extracted all the sentences with phrasal verbs. Then, eight linguistic factors were encoded into each extracted sentence, and the annotated data were statistically analyzed with GLM.

Through the analysis, it was demonstrated that Korean EFL learners employed a different strategy in the particle placement and that only some factors were used for the selection of constructions. Unlike native speakers, only four linguistic factors were significant in Korean EFL learners' writings (ANIMACY, PRONOMINALITY, CONCRETENESS, and LENGTH). It was also observed that there were some differences in the ratio of these two constructions (*construction_0* vs. *construction_1*) as the level of proficiency went up.

Lee et al. (2016) extended the scopes of study and statistically examined the particle placement of the EFL learners' writings in three East Asian countries (Chinese, Japanese, and Korean). They manually encoded eleven linguistic factors (Table 4) and statistically analyzed the data with R. The study also adopted a GLM analysis and statistically analyzed how each factor influenced the choice of alternation. The study also took the ICE-GB corpus as a reference corpus and compared the tendencies of the three EFL learners' writings with those of native speakers.

3 Research Method

3.1 Corpus

This study employed two types of data. The first one was the TOEFL11 corpus for the EFL learners (LDC Catalo No.: LDC2014T06), and the second one was the data in ICE-GB for the native speakers (as reference data set).

The TOEFL11 corpus was first released by the English Testing Service (ETS) in 2014. The corpus consists of essays written during the TOEFL iBT® tests in 2006-2007 (Blanchard et al., 2013). It contains 1,100 essays per each of the 11 native languages, totaling 12,100 essays. All of the essays were taken from the parts of TOEFL independent task, where test-takers were asked to write an essay in response to a brief writing topic. The essays were sampled as evenly as possible from eight different topics. The corpus also provides the score levels (Low/Medium/High) for each essay.

From the TOEFL corpus, three components were chosen: Chinese, Japanese, and Korean. The sizes of each component were as follows.

Level	Chinese	Japanese	Korean
High	102,293	67,404	95,066
Medium	228,331	194,716	202,531
Low	21,798	40,060	30,787
Total	352,422	302,180	328,384

Table 3: Corpus Size of Each Component

These texts were the target of the investigations.

The ICE-GB corpus contained both spoken and written components of native speakers in Great Britain. Its size was about 1 million (word) tokens. Among the corpus data in this corpus, only the written part was taken, since the data for three EFL corpora were written materials.

3.2 Procedure

The analysis in this paper proceeded as follows.

First, four corpora were chosen for the analysis: Chinese, Japanese, Korean and ICE-GB.

Second, each text in the three EFL corpora was POS tagged with the C7 CLAWS taggers.[2]

---

[2] You can easily use Free CLAWS WWW tagger in http://ucrel.lancs.ac.uk/claws/trial.html. For details of C7 tag sets, see Jurafsky and Martin (2009).

Third, all the sentences with particles were extracted using NLPTools (Lee, 2007).[3]

Fourth, since the number of extracted sentences in three EFL learners' corpora was about 1,000, exactly 1,000 sentences were randomly extracted from the ICE-GB corpus.

Fifth, eleven linguistic factors were encoded to each sentence.[4] They are enumerated in Table 4.[5]

Tag Type	ID Tag	ID Tag Level
Length	LENGTHS	
	LENGTHW	
Synatx	VOICE	active, passive
	NPTYPE	proper noun, lexical, semi-pronominal, pronominal
	DEFINITE	definite, indefinite, no determiner
	COMPLEX	simple, intermediate, complex
	PP	yes, no
	PART=PP	yes, no
Semantics	ANIMACY	animate, inanimate
	IDIOMACITY	literal, metaphorical, idiomatic
	CONCRETENESS	abstract, concrete

Table 4. Variables Used in the Analysis

Following the study of Atkins (1987), each linguistic factor and its level were called ID tag and ID tag levels respectively. These variables were used in the statistical analysis.

Finally, all the data were statistically analyzed using R.

4 Correlation Analysis

4.1 Preprocess

After all the sentences with the particles were extracted from each corpus and eleven linguistic factors were encoded to the extracted sentences, all the data were statistically analyzed using R (R

---

[3] In the C7 tag sets, particles have a tag RP. The reason why NLPTools was used here is that the software had a function which could extract the whole sentences with the given tag(s) (i.e., *_RP).

[4] This operation is called *operationalization* (Deshors, 2010, Deshors and Gries, 2014).

[5] As you can find in this table, all the factors which were related with the discourse properties were not included in the encoding process.

Core Team, 2016). Before the statistical analyses were performed based on Gries (2013) and Lee (2016), the sentences were classified into two groups, based on the transitive vs. intransitive use of phrasal verbs. This process was necessary since the particle placement occurred only in the transitive or ditransitive use of phrasal verbs.

The first statistical analysis which was taken was the correlational analysis. This paper followed the correlation analysis in Gries (2001:41), and the coefficients in Table 5 were determined depending on the measurement scale of the variables.

Variable	Correlation Coefficient
Categorical	ϕ, Cramer's V, and λ
Ordinal	γ (equaling Kendall's τ with correction for ties)
Interval	Pearson product-moment correlation

Table 5: Coefficients for Each Measurement Scale

These (monofactorial) correlation analyses were only taken in order to numerically examine how each linguistic factor influenced the choice of alternation.

Though this paper followed the analysis in Gries (2001), there were three differences between the analysis in Gries (2001) and those of this paper. First, though Gries (2001) provided the correlation coefficients for both each linguistic factor and each value for each linguistic factor, this paper provided the correlation coefficients only for each linguistic factor (not each value for the factor). This strategy was chosen since we were primarily interest in how each linguistic factor influenced the choice of alternation, not how the value for each linguistic factor was.[6] Second, Gries (2001) used Crammer's *I*, but this paper used Crammer's *V* instead. This difference was originated from the above strategy. Since Gries (2001) provided the coefficients for both each linguistic factor and each value for each linguistic factor, nominal data prevailed. Since this paper provided the correlation coefficients only for each linguistic factor (not each value for the factor), categorical data were abundant. Thus, Crammer's *V* was more appropriate in this paper. Third, the (correlation) coefficients for native speakers were provided for the data which were collected from the BNC corpus. However, this paper employed

[6] You might investigate the value for each linguistic factor in the analysis of Lee et al. (2016).

the ICE-GB corpus. In addition, as mentioned in Section 3.2, only 1,000 sentences were randomly extracted from the ICE-GB corpus. Accordingly, a comparison with Gries (2001) was impossible.

The following four tables illustrated the analysis results in the ICE-GB corpus and those of three components in the TOEFL corpus.

Variables	Correlation Coefficient
LENGTHS	r_{pbis}=0.587 ***
LENGTHW	r_{pbis}=0.542 ***
VOICE	ϕ=0.244 * (λ=0)
NPTYPE	γ=-0.819 ***
DEFINITE	V=0.215 ***
COMPLEX	γ=0.717 ***
PP	ϕ=0.1 * (λ=0)
PART=PP	ϕ=0.021 ** (λ=0)
ANIMACY	ϕ=0.3 *** (λ=0)
IDIOMACITY	V=-0.03 ***
CONCRETENESS	ϕ=-0.36 *** (λ=0.259)

Table 6: Correlation Analysis (English)

Variables	Correlation Coefficient
LENGTHS	r_{pbis}=0.684 ***
LENGTHW	r_{pbis}=0.645 ***
VOICE	ϕ=0.12 * (λ=0)
NPTYPE	γ=-0.912 ***
DEFINITE	V=0.169 ***
COMPLEX	γ=0.826 ***
PP	ϕ=0.07 * (λ=0)
PART=PP	ϕ=0.28 ** (λ=0)
ANIMACY	ϕ=0.017 *** (λ=0)
IDIOMACITY	V=-0.09 ***
CONCRETENESS	ϕ=-0.33 *** (λ=0.218)

Table 7: Correlation Analysis (Chinese)

Variables	Correlation Coefficient
LENGTHS	r_{pbis}=0.784 ***
LENGTHW	r_{pbis}=0.713 ***
VOICE	ϕ=0.068 * (λ=0)
NPTYPE	γ=-0.968 ***
DEFINITE	V=0.187 ***
COMPLEX	γ=0.933 ***
PP	ϕ=0.08 * (λ=0)
PART=PP	ϕ=0.015 ** (λ=0)
ANIMACY	ϕ=0.15 *** (λ=0)
IDIOMACITY	V=-0.15 ***
CONCRETENESS	ϕ=-0.32 *** (λ=0.174)

Table 8: Correlation Analysis (Korean)

Variables	Correlation Coefficient
LENGTHS	r_{pbis}=0.636 ***
LENGTHW	r_{pbis}=0.577 ***
VOICE	ϕ=0.066 * (λ=0)
NPTYPE	γ=-0.938 ***
DEFINITE	V=0.066 ***
COMPLEX	γ=0.788 ***
PP	ϕ=0.12 * (λ=0)
PART=PP	ϕ=0.013 ** (λ=0)
ANIMACY	ϕ=0.28 *** (λ=0)
IDIOMACITY	V=0.07 ***
CONCRETENESS	ϕ=-0.23 *** (λ=0.126)

Table 9: Correlation Analysis (Japanese)

These tables demonstrated that each linguistic factor played a role differently in each variety of English.

In these four tables, the following three facts were observed. First, the absolute values for the coefficients in two factors (NPTYPE and COMPLEX) were over 0.7. We usually say that the relationship is strong if the coefficient is over 0.7. Accordingly, we could say that the relationship was strong in these two factors. Second, the absolute values for the coefficients in two factors (LENGTHS and LENGTHW) were between 0.3 and 0.7. We usually say that the relationship is moderate if the value is between 0.3 and 0.7. Consequently, we could say that the relationship was moderate in these two factors. Third, the values for the coefficients in the other factors were under 0.3. We usually say that the relationship is weak if the coefficient is under 0.3. Thus, we could say that the relationships were weak in the other factors.

A close comparison of these tables revealed (i) that the values of native speakers (English) were similar to those of Japanese EFL learners and (ii) that the values of Chinese EFL learners were similar to those of Korean EFL learners.

5 Agglomerative Clustering

In order to examine which one was close to which one, another statistical analysis was performed. The second statistical analysis was a hierarchical agglomerative clustering analysis.

Usually, the cluster analyses have been used to determine the similarity among the group members or the degree of granularity exhibited by the group members. In this paper, the tables of correlation coefficients were submitted into a hierarchical agglomerative cluster analysis, resulting in the dendrogram in the following figure.
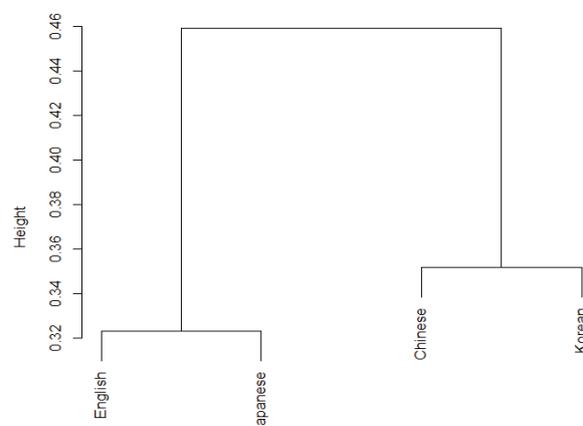

Figure 1: Cluster Dendrogram

Here, the horizontal lines indicate which one can be grouped with which one, and the vertical lines indicate the distance between the two groups.

This figure graphically illustrated the following facts. First, the correlation relationships of native speakers (i.e., English) were similar to those of Japanese EFL learners, which can be represented as {English, Japanese}. Second, the relationships of Chinese EFL learners were similar to those of Korean EFL learners, which can be represented as {Chinese, Korean}. Third, these two groups were amalgamated together to form a cluster {{English, Japanese}, {Chinese, Korean}}. The dendrogram in Figure 1 demonstrated (i) that the tendency of particle placement of Japanese EFL learners was close to that of native speakers and (ii) that the tendencies of particle placement of Chinese and Korean EFL learners were slightly far from that of native speakers.

6 Discussions

In this paper, the alternation of particle placement was closely examined in the native speakers' writings and the three EFL learners' writings.

From the three components of the TOEFL11 corpus and the ICE-GB corpus, all the sentences with phrasal verbs were extracted and eleven factors were manually encoded into the extracted sentences.

The correlation analyses between these eleven linguistic factors and the choice of alternation revealed the following facts. First, the coefficients

in two linguistic factors (NPTYPE and NPTYPE) were strong since the coefficient is over 0.7. This means that the data points for these two linguistic factors were closely distributed to the regression lines for these factors. Second, the coefficients in two linguistic factors (LENGTHS and LENGTHW) were moderate since the coefficient was between 0.3 and 0.7. This means that the data points for these linguistic factors were moderately distributed to the regression lines for these factors. Of course, the coefficients of LENGTHS and LENGTHW were over 0.7 in Korean. This says that the relationship of these factors was strong in the Korean EFL learners' writings. Third, the coefficients in the other linguistic factors were weak since the values were under 0.3. This means that the data points for these linguistic factors were sparsely distributed to the regression lines for these factors.

Based on the results of correlation analysis, a hierarchical agglomerative clustering analysis was performed. This analysis was conducted in order to examine which one was close to which one. The analysis result was {{English, Japanese}, {Chinese, Korean}}, where the correlation relationships of native speakers (i.e., English) were similar to those of Japanese EFL learners and the relationships of Chinese EFL learners were similar to those of Korean EFL learners.

If it had been supposed that the EFL learners were severely influenced by the L1, the correlation analysis results would have been different. That is, if it had been supposed that the L1 transfer effects had involved in the choice of particle placement in the three EFL learners' writings (following the study of Oldin [1989, 2003]), the analysis result would have been {{English}, {Japanese, Chinese, Korean}}. The dendrogram in Figure 1 illustrated that more factors might be involved in the choice of particle placement constructions in the Japanese EFL learners, as Lee et al. (2016) mentioned. More study is necessary to investigate what linguistic or extra-linguistic factors influenced this kind of tendency.

7   Conclusion

This paper adopted a monofactorial analysis as in Gries (2001, 2003) to examine particle placement in three East Asian EFL learners' writings. For the comparison, two different types of corpora were chosen. The components of the TOEFL11 corpus was used for the EFL learners' parts (Chinese, Japanese, and Korean), and the ICE-GB corpus was chosen for the native speakers' parts. Then, all the relevant sentences were extracted using the tag information. After that, the eleven relevant factors were encoded to these sentences, and each factor and their interactions were statistically analyzed with R.

Through the correlation analysis, it was found that each linguistic factor influenced differently in four varieties of English. Through the hierarchical agglomerative clustering analysis, it was found that the correlation relationships of native speakers (i.e., English) were similar to those of Japanese EFL learners and the relationships of Chinese EFL learners were similar to those of Korean EFL learners.

However, we do NOT say that these differences between the native speakers and the three East Asian EFL learners come from only the L1 transfer effects. Another kind of complicated statistical analysis (such as another regression analysis with the native data and/or the analysis in Gries and Deshors (2015)) is necessary to examine if the L1 really influenced these factors and how much the L1 transfer effects are involved in these factors. Notwithstanding, this mentioning does not say that the analysis result in this paper is meaningless. The analysis results in this paper enumerate how much each linguistic factor influenced the choice of particle placement, and we can start our future research from this set of factors.

References

Bernhard Rohrbacher. 1994. English Main Verbs Move Never. The Penn Review of Linguistics, 18:145-159.

Beryl Atkins. 1987. Semantic ID Tags: Corpus Evidence for Dictionary Senses. In Proceedings of the Annual Conference of the UW Center for the New Oxford English Dictionary, 17-36. University of Waterloo, Waterloo, ON, Canada.

Bruce Fraser. 1974. The Phrasal Verb in English, by Dwight Bolinger. Language, 50:568-575.

Bruce Fraser. 1976. The Verb-Particle Combination in English. New York: Academic Press.

Daniel Blanchard, Joel Tetreault, Derrick Higgins, Aoife Cahill, and Martin Chodorow. 2013. TOEFL11: A corpus of non-native English. ETS RR–13-24. Prinston, NJ: Educational Testing Service.

Daniel jurafsky and James Martin. 2009. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. Prentice Hall, Upper Saddle Hill, NJ.

Etsko Kruisinga and Patrick Erades. 1953. An English Grammar. Vol. I. P. Noordhoff, Groningen.

Etsko Kruisinga and Patrick Erades. 1953. An English Grammar. Vol. I. Groningen: P. Noordhoff.

Gerald Nelson, Sean Wallis, and Bas Aarts. 2002. Exploring Natural Language: Working with the British Component of the International Corpus of English. John Benjamins Publishing Company, Amsterdam.

Guy Aston and Lou Burnard. 1998. The BNC Handbook. Exploring the British National Corpus with SARA. Edinburgh University Press, Edinburgh.

Gyu-hyoeng Lee, Ha-Eung Kim, and Yong-hun Lee. 2015. A Multifactorial Analysis of English Particle Movement in Korean EFL Learners' Writings. Proceedings of 19th Pacific Asian Conference on Language, Information, and Computation. Shanghai, China.

Henry Sweet. 1892. A New English Grammar. Clarendon Press, Oxford.

Henry Sweet. 1892. A New English Grammar. Oxford: Clarendon Press.

John Hawkins. 1994. A Performance Theory of Order and Constituency. Cambridge: Cambridge University Press.

Marcel Den Dikken. 1995. Particles: On the Syntax of Verb-Particle, Triadic, and Causative Constructions. Oxford: Oxford University Press.

Marcel Den Kikken. 1992. Particles. Holland Institute of Linguistics Dissertations. The Hague: Holland Academic Graphics.

Noam Chomsky. 1957. Syntactic Structures. Mouton, Berlin.

Noam Chomsky. 1965. Aspects of the Theory of Syntax. MIT Press. Cambridge.

Otto Jespersen. 1928. A Modern English Grammar on Historical Principles. George Allen and Unwin Ltd., London.

Otto Jespersen. 1928. A Modern English Grammar on Historical Principles. London: George Allen and Unwin Ltd.

Ping Chen. 1986. Discourse and Particle Movement in English. Studies in Language 10:79-95.

R Core Team. 2016. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing.

Rosemary Yeagle. 1983. The Syntax and Semantics of English Verb-Particle Constructions with off: A Space Grammar Analysis. Unpublished M.A. Thesis, Southern Illinois University at Carbondale

Sandra Deshors and Stefan Th. Gries. 2014 A Case for the Multifactorial Assessment of Learner Language: The Use of *May* and *Can* in French-English Interlanguage. In Dylan Glynn and Justyna Robinson (eds.), Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy, 179-204. John Benjamins Publishing Company, Amsterdam.

Sandra Deshors. 2010. Multifactorial Study of the Use of *May* and *Can* in French-English Interlanguage. Ph.D. dissertation. University of Sussex.

Stefan Th. Gries and Sandra Deshors. 2015. EFL and/vs. ESL? A Multi-level Regression Modeling Perspective on Bridging the Paradigm Gap. International Journal of Learner Corpus Research 1(1): 130–159.

Stefan Th. Gries. 1999. Particle movement: A Cognitive and Functional Approach, Cognitive Linguistics, 10(2):105-145.

Stefan Th. Gries. 2001. A Multifactorial Analysis of Syntactic Variation: Particle Movement Revisited. Journal of Quantitative Linguistics, 8(1):33-50.

Stefan Th. Gries. 2003. Multifactorial Analysis in Corpus Linguistics: A Study of Particle Movement. Continumm, London.

Stefan Th. Gries. 2013. Statistics for Linguistics with R: A Practical Introduction. Guyter, Berlin.

Terence Oldin. 1989. Language Transfer. Cambridge University Press, Cambridge.

Terence Oldin. 2003. Cross-linguistic Inference. In Catherine Doughty and Michael Long (eds.), The Handbook of Second Language Acquisition, 436-486. Blackwell, Oxford.

Yong-hun Lee, Ha-Eung Kim, and Gyu-hyoeng Lee. 2016. A Multifactorial Analysis of English Particle Placement in Three East Asian Countries. A Submitted Paper.

Yong-hun Lee. 2007. Corpus Analysis Using NLPTools and Their Applications: Applications to Linguistic Research, English Education, and Textbook Evaluation. Cambridge University Press, Seoul.

Yong-hun Lee. 2016. Corpus Linguistics and Statistics Using R. Hankook Publishing Company, Seoul.